

MÉTODO PARA LA DETERMINACIÓN OBJETIVA DE ERRORES DE POSICIÓN EN PREDICCIONES DE LLUVIA

José Antonio López Díaz⁽¹⁾, Carlos Geijo Guerrero⁽²⁾

⁽¹⁾ AEMET, jlopezd@aemet.es

⁽²⁾ AEMET, cgeijog@aemet.es

Resumen

En este trabajo se presenta un método objetivo para la valoración de los errores de posición en las predicciones de lluvia a resolución mesoscalar. A esta escala es habitual la formación de varios clusters de precipitación alrededor de sus centros y surge por tanto la cuestión de cómo medir de una forma objetiva la disparidad de las posiciones de los clusters entre el modelo y un campo verificador. El enfoque para la verificación que adoptamos aquí atiende por tanto a la estructura de la precipitación más que a medidas basadas en estadísticos sobre la rejilla, que han sido los más habituales hasta hace poco. Esta técnica ha sido aplicada con éxito a una serie de predicciones elaboradas con el modelo numérico de predicción a mesoescala (2.5Km) HARMONIE. Como dato de verificación se ha empleado el dato radar de uno de los radares de la red operativa de AEMET. Los resultados muestran que el algoritmo es robusto y que puede utilizarse en modo “bootstrapping” con un alto número de simulaciones aleatorias para obtener una medida objetiva de la pericia del modelo en ubicar zonas de precipitación.

1. Metodología

1.1 Método de verificación de errores en la posición de clusters (VPC)

Se presenta en este trabajo una metodología novedosa para comparar las posiciones de dos conjuntos de clusters y medir el grado de discrepancia entre las mismas. Esta metodología tiene la potencialidad de permitir el planteamiento de un test estadístico objetivo para valorar el grado de discrepancia. Por otra parte el enfoque adoptado en el VPC atiende a la estructura de la precipitación más que a medidas basadas en estadísticos sobre la rejilla, que han sido los más habituales hasta hace poco. En este sentido tiene similitudes con el método SAL de verificación, procedimiento que se fija en la estructura (Structure), amplitud (Amplitude) y posición (Location) de los campos de lluvia.

Sin embargo el SAL mira a aspectos globales del campo de lluvia mientras que el método que aquí proponemos hace un detallado análisis de las posiciones de los clusters de precipitación.

1.1.1 Descripción del método: grado de disparidad

La base del método VPC consiste en definir una magnitud, el grado de disparidad, que sintetiza de forma eficiente la medida en que los dos conjuntos de clusters que se comparan ocupan posiciones distintas dentro de una zona de estudio.

Para definir el grado de disparidad entre los dos conjuntos de posiciones de centros de clusters, que pueden tener distinto número de elementos en el modelo y en las observaciones, se ha utilizado un emparejamiento óptimo entre los dos conjuntos. Si los centros de los N clusters del modelo se designan con y_1, y_2, \dots, y_N , y los centros de los M clusters de las observaciones por x_1, x_2, \dots, x_M , el emparejamiento óptimo utilizado se consigue seleccionando un conjunto de líneas que unen puntos y_i con puntos x_j , satisfaciendo las tres condiciones siguientes:

- a) sale al menos una línea de cada punto y_i , $i = 1, 2, \dots, N$
- b) llega al menos una línea a cada punto x_j , $j = 1, 2, \dots, M$
- c) la suma de las longitudes de todas las líneas es mínima.

Con este procedimiento de emparejado se asegura que todos los centros de clusters del modelo están asociados al menos a un centro de cluster de observaciones y viceversa. Además la tercera condición asegura que el emparejamiento de centros es óptimo, de tal manera que no es posible emparejar completamente los dos conjuntos de clusters consumiendo un total de longitudes de líneas menor. De esta forma la suma de longitudes de las líneas del emparejamiento óptimo es una medida idónea de la discrepancia en las posiciones de los centros. Esta suma de longitudes la denominaremos en lo sucesivo *grado de disparidad (GD)*.

1.1.2 Algoritmo aproximado

Así planteado el problema de emparejamiento óptimo resulta muy complejo. Se ha conseguido sin embargo un algoritmo relativamente sencillo que permite encontrar una solución rápidamente. Esta solución proporciona una suma de longitudes en algunos casos óptima y en otros muy próxima a la óptima, con lo cual permite la aplicación práctica de esta técnica (ver 1.1.3 abajo). Este algoritmo efectúa, en un primer paso, una búsqueda por el conjunto de

las $N \times M$ líneas entre centros de modelo y observaciones, ordenadas de menor a mayor longitud, seleccionando en cada paso una nueva línea si esta une algún centro de modelo u observaciones sin línea. Se termina cuando todos los centros de modelo y observaciones tienen al menos una línea. En un segundo paso se eliminan líneas redundantes, es decir, líneas que se pueden eliminar sin violar las dos primeras condiciones de antes, eliminando primero las más largas.

Un ejemplo en que la solución proporcionada por el anterior algoritmo no es la óptima se presenta en la figura 1.

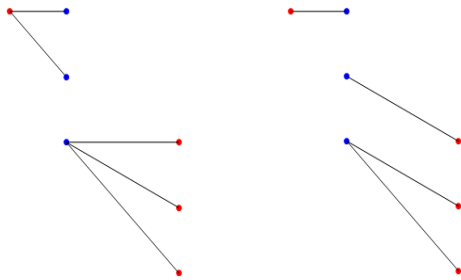


Fig. 1: En rojo posición de los clústeres de observaciones, en azul de los modelos. Izquierda solución con el algoritmo aproximado, derecha solución exacta.

Vemos que la suma de longitudes, o GD, en la solución con el algoritmo aproximado de la izquierda es superior al grado de disparidad de la situación de la derecha, que es el correcto según la definición.

1.1.3 Validación del algoritmo aproximado.

N	4	7	10
M	3	4	10
% DIF	45.7	41.2	0.2
E.Rel(%) Med	11.7	6.0	11.1

Tabla 1: Resultados de la comparación del algoritmo aproximado con la solución óptima. Explicación en texto.

Al objeto de evaluar la bondad del algoritmo aproximado para el VPC se han efectuado simulaciones con diferentes combinaciones de números de clústeres del modelo N y de clústeres de las observaciones M . Para cada combinación de N y M se han simulado 1000 casos con posiciones aleatorias para los puntos tanto del modelo como de las observaciones. Para cada caso simulado se ha aplicado el algoritmo aproximado y también se ha buscado la solución óptima para el VPC con un

algoritmo de búsqueda exhaustiva que garantiza la solución óptima. El algoritmo de búsqueda exhaustiva se basa en una función recursiva que va añadiendo en cada paso una nueva línea entre las disponibles restantes que unen un punto del modelo y uno de las observaciones, cortando si no quedan suficientes líneas para completar el grafo o si se excede el límite de suma de longitudes de líneas ya hallado en algún grafo completo anterior.

Los resultados se presentan en la tabla 1. La fila % DIF da el porcentaje de casos en que hay diferencia entre ambos procedimientos. La última fila es la media de los errores relativos en las grado de disparidad en los casos en que hay diferencia. Vemos que las diferencias son mayores para los valores más pequeños de N y M . Pero incluso para el caso de valores más pequeños, $N = 4$ y $M = 3$, vemos que en más del 50% de los casos el algoritmo aproximado consigue la solución óptima, y el porcentaje medio de error global del algoritmo aproximado está alrededor del 5%. En cambio en el caso de mayor número de puntos N y M el algoritmo aproximado solo difiere de la solución óptima en 2 de cada 1000 simulaciones. Conviene recordar que en casos como este, con altos números de puntos, es donde el algoritmo de búsqueda exhaustiva es comparativamente más costoso computacionalmente.

1.1.4 Test estadístico basado en el grado de disparidad.

Ya se comentó antes que una de las ventajas del método VPC es que permite el planteamiento de un test de hipótesis estadístico. En concreto, podemos plantear un test estadístico con las siguientes hipótesis nula y alternativa:

H_0 : las posiciones de los dos conjuntos de clústeres (modelo y observaciones) se distribuyen aleatoriamente en una región dada (por ejemplo un rectángulo)

H_1 : los dos conjuntos de posiciones muestran un grado de disparidad inferior al aleatorio (o sea, están más próximos que en condiciones aleatorias)

Para el cálculo del p-valor del anterior test tendremos que realizar simulaciones aleatorias (técnica de Monte Carlo) dado que la distribución exacta del grado de disparidad bajo la hipótesis nula es desconocida. Seguiremos los siguientes pasos:

- ✓ Computamos con el algoritmo aproximado descrito el grado de disparidad Δ entre los clústeres del modelo y los de las observaciones
- ✓ Simulamos N posiciones aleatorias para los clústeres del modelo

- ✓ Para estas posiciones computamos el valor de GD
- ✓ Efectuamos los dos pasos anteriores un número grande (por ejemplo 5000) de veces.
- ✓ Ordenamos de menor a mayor los valores de grado de disparidad obtenidos: el p-valor del test es la posición de Δ en la serie ordenada (expresada en % respecto al n° total de simulaciones)

De esta forma obtenemos una medida objetiva del grado de discrepancia entre las posiciones de los centros de clústeres del modelo y las observaciones, lo cual indudablemente es una ventaja frente a técnicas de verificación tipo SAL que no permiten el cálculo de p-valores.

1.1.5 Problemática derivada del tamaño de los clústeres.

En una aplicación real de la VPC con el cálculo del p-valor descrito en el apartado anterior, los clústeres tienen un tamaño no nulo. Si queremos respetar al máximo la estructura de los campos del modelo y observados, debemos evitar que en las simulaciones los clústeres se solapen. Una posible solución consiste en que en cada simulación aleatoria de posiciones asignemos a cada posición aleatoria simulada un clúster y sólo aceptemos esa posición si poniendo ahí el clúster, este no solapa con ninguno de los anteriores. Continuamos buscando posiciones hasta asignar posiciones aceptables a todos los clústeres. De esta forma se garantiza la no superposición de clústeres, pero a cambio se restringe la libertad en la elección de posiciones nuevas.

Esta restricción puede dar lugar a un problema si los tamaños de los clústeres son grandes con relación al área de la región de estudio: las posiciones aleatorias podrán variar relativamente poco y por tanto la dispersión de valores de GD simulados será pequeña. En consecuencia el p-valor obtenido en la aplicación del test puede ser muy pequeño (próximo a cero). Lo que sucede en esta situación es que disponemos de una referencia escasa con la que valorar el grado de disparidad obtenido. Cabe concluir de este análisis que no conviene, para la correcta aplicación de la técnica VPC aquí descrita, que los clústeres de modelo o de observaciones ocupen un espacio grande del área de estudio.

2. Aplicación del Método

Este método de verificación de errores de posición de clusters ha sido aplicado en un estudio de impacto de la asimilación de dato radar en el modelo NWP HARMONIE (<http://hirlam.org/index.php>). Se trata de un modelo de predicción no hidrostático de alta resolución espacial (2.5Km, 60 niveles) que permite una representación explícita (i.e., no parametrizada) de la convección. El algoritmo de asimilación empleado es 3DVar y el módulo de asimilación de dato radar de reflectividad y de viento Doppler ha sido desarrollado por Météo-France.

En el experimento se utilizó el dato procedente del C-radar de la red operativa de AEMET ubicado en las proximidades de Madrid, en concreto se utilizaron datos captados durante un episodio de lluvia de tipo frontal que tuvo lugar entre los días 04/11/2011 a las 11UTC y el 05/11/2012 a las 20UTC. Aunque hay dato disponible cada diez minutos, en la verificación se empleó únicamente el dato con frecuencia horaria. Se consideraron las 2 elevaciones más bajas (1.4 y 0.5 grados sobre la horizontal) de las 13 elevaciones de pulso corto generadas por el radar. Aunque puede parecer que esta elección supone aprovechar solo una pequeña fracción del total de información disponible, esto no es así debido a que para elevaciones mayores el recorrido del haz en la troposfera se acorta significativamente y por tanto su cobertura se limita a una pequeña área alrededor del radar. Aunque estas elevaciones pueden sin duda tener interés en casos de convección profunda, se dejaron fuera en este estudio en aras de una mayor facilidad en el manejo de los datos.

El experimento en sí consistió en la ejecución con HARMONIE de una serie de predicciones cada 3 horas (00, 03,...21) con un alcance de 18 horas cada una. Esto corresponde a 11 predicciones durante el episodio elegido. Junto con el conjunto de predicciones de “control”, se hicieron pruebas con tres variantes en la especificación de las condiciones iniciales, lo que arroja un total de 44 predicciones. Para la determinación de estas condiciones iniciales se utilizaron como “first-guess” los campos del modelo ECMWF y como observaciones únicamente los datos radar de reflectividad y viento radial, es decir, se trata de una configuración diferente a la habitual tanto en lo referente al carácter no “autoregresivo” del proceso de predicción, como en lo referente al uso limitado de observaciones.

2.1 Preparación de la verificación

Las predicciones a verificar fueron empleadas para generar “pseudo-imágenes radar” mediante un simulador de reflectividades incorporado en el sistema HARMONIE. Estas pseudo-imágenes son

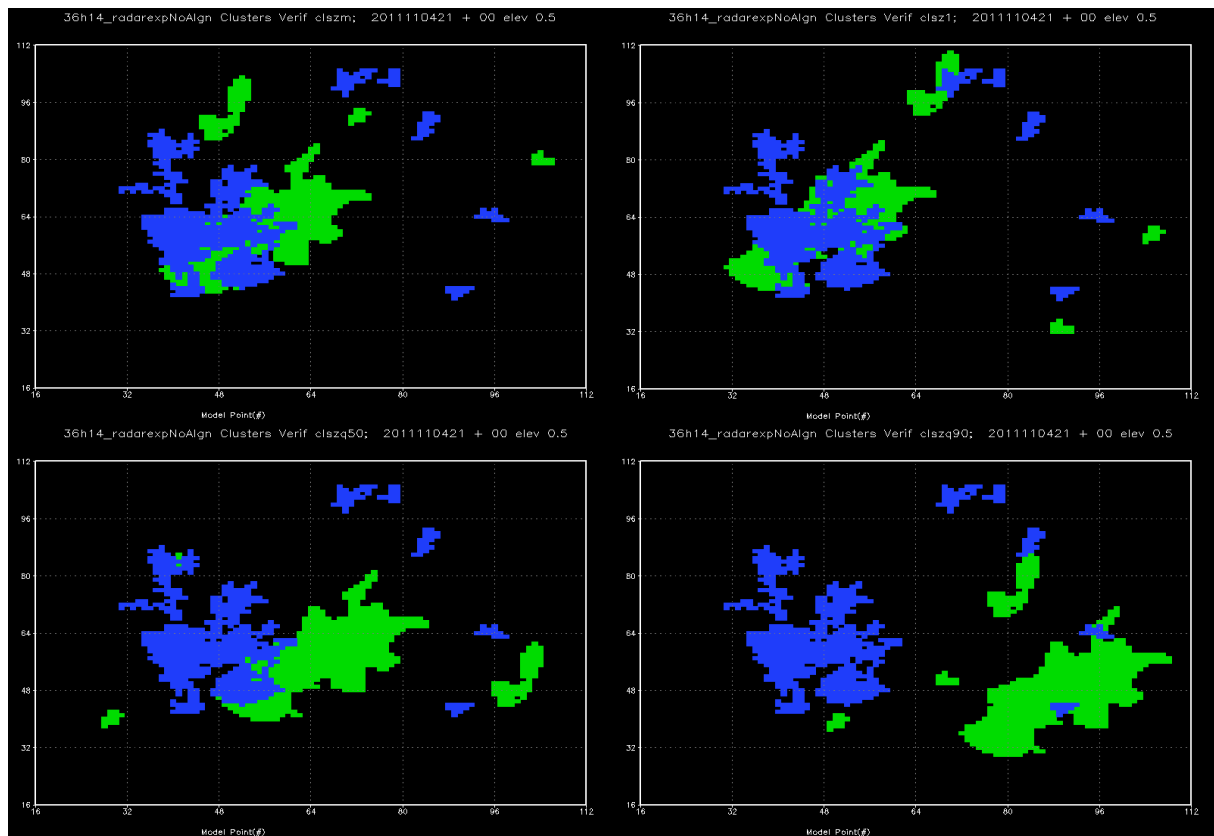


Fig. 2: Ejemplo de determinación del p-valor para los errores de posición mediante muestreo aleatorio. Los clusters en azul corresponden al campo de reflectividad detectado por el radar (4/11/2011 a las 21UTC, elevación 0.5 grados) “up-scaled” a la resolución del modelo. Clusters en verde corresponden a “pseudo-imágenes” de un análisis HARMONIE para el mismo instante, y en geometría radar. El cuadro superior izquierdo del panel muestra la situación a verificar. A la derecha se presenta el caso con mínimo valor de GD generado aleatoriamente en una muestra de tamaño 1000. En la fila inferior tenemos la mediana y el percentil 90. El caso a verificar obtiene un p-valor de 12.4

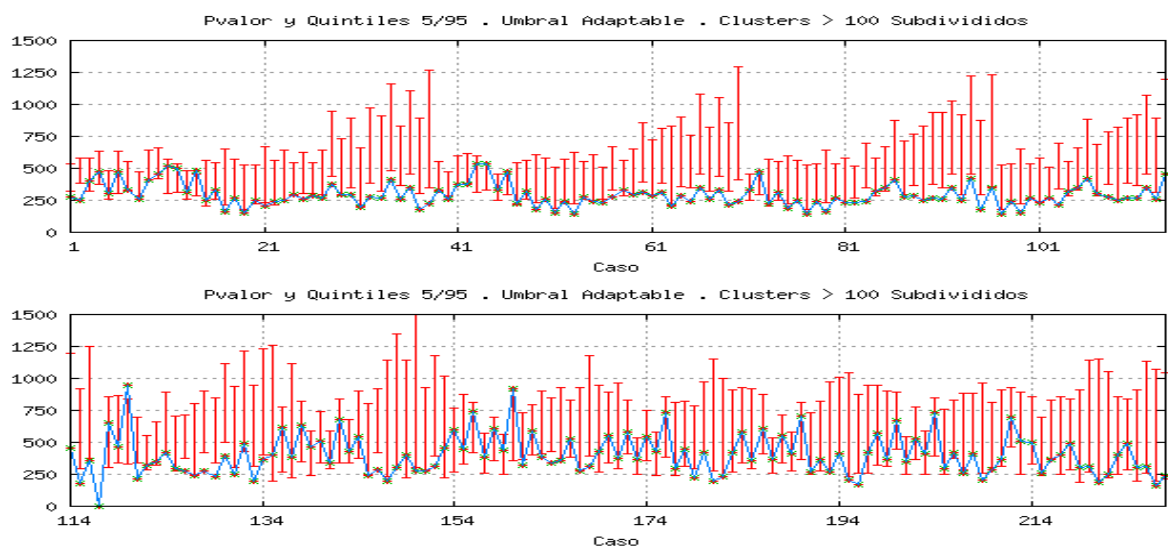


Fig. 3: Serie completa de GD (línea azul) y percentiles 5% (extremo inferior de las barras verticales) y 95% (extremo superior) de los valores GD simulados para uno de los experimentos realizados. Ver texto para más detalles

producidas a una resolución y con una geometría diferente a las imágenes suministradas por el radar y deben por tanto ser sometidas a un postproceso que permita hacer una comparación homogénea. En este caso se optó por reducir la resolución espacial del dato radar, que en la dirección radial es de 500m, a la resolución del modelo que es de unos 2500m.

A continuación se procede a delimitar los clusters que precisa el método de verificación. Para ello se ha empleado un viejo algoritmo de “4-conectividad” (González y Woods, 1992) que agrupa pixels con sus vecinos laterales si estos satisfacen una condición dada. La condición empleada en este caso es simplemente que el valor de reflectividad (o pseudo-reflectividad) esté por encima de un cierto umbral, por ejemplo 15 dBZ. Como se ha mencionado más arriba, el proceso de verificación resultó fuertemente dependiente de la elección de este umbral. La razón de ello reside en que la diferencia de tamaños entre clusters observados y simulados resultó ser importante. Esta diferencia es en sí misma interesante y puede ser debida a una serie de razones que merecen sin duda un estudio detallado, pero desde el punto de vista del método de verificación de errores de posición aquí presentado supone una serie de inconvenientes.

Por un lado, la posición de un cluster se define inicialmente en términos de su centro de masa. Esta definición resulta insuficiente cuando el cluster, debido a su tamaño, se diferencia ostensiblemente de un punto. Por ejemplo, no pueden tenerse en cuenta efectos de solapamiento. También en esta situación la diferencia entre el número de clusters verificados (radar) y verificados (modelo) suele ser importante porque el cluster del modelo absorbe la mayor parte de los pixels en la escena que satisfacen la condición umbral. Según la tabla 1, el método aproximado empeora sus resultados cuando la disparidad en el número de clusters observados y simulados crece. Finalmente, como se ha explicado en 1.1.5, también el proceso de muestreo por “bootstrapping”, a través del cual se obtiene una medida objetiva de los errores de posición, resulta perjudicado por estos clusters desproporcionadamente grandes.

Después de algunas pruebas, resultó evidente que los resultados del método mejoran cuando se adopta una doble solución. Por una parte, en lugar de utilizar un mismo umbral para observaciones y modelo, se utilizan umbrales adaptables a la escena, los cuales garantizan que el porcentaje total de clusters agrupados en observaciones y modelo son similares. Este procedimiento permite soslayar de una forma sencilla y eficaz los problemas de sesgos entre observaciones y modelo. Por otra parte, clusters cuyo tamaño excede un valor dado, son subdivididos, de tal suerte que finalmente tenemos

una colección de clusters de tamaños comparables. De esta manera el método captura mucho mejor efectos de forma y solapamiento.

El método ha sido implementado en FORTRAN-95 haciendo uso de técnicas próximas a la programación orientada a objetos (tipos derivados de datos, sobrecarga de procedimientos) que son posibles en este lenguaje de programación y que resultan particularmente convenientes en este caso, ya que toda la arquitectura de la utilidad gira entorno a objetos como “cluster” y “flecha”.

2.2 Resultados de la verificación

La figura 2 muestra un ejemplo de determinación del p-valor para una situación dada. Obsérvese que los clusters del modelo (verde) han sido desplazados aleatoriamente desde su posición original (arriba a la izquierda) de forma paralela y evitando solapes. En este caso, el modelo es batido por el proceso aleatorio un 12% de las veces.

La figura 3 muestra la serie completa de valores de GD en ordenadas (línea azul con estrellas) y los percentiles 5% (extremo inferior de las barras verticales), y 95% (extremo superior) de los valores de GD simulados aleatoriamente para uno de los experimentos realizados. Por tanto, si el valor de GD está por debajo del extremo inferior de la barra roja quiere decir que el test sobre el GD es significativo al 5%. La evidente periodicidad que se aprecia en la figura es consecuencia de la disposición consecutiva a lo largo de las abscisas de los análisis y las predicciones de cada uno de los 11 ciclos a lo largo del periodo elegido. Los resultados para las elevaciones 0.5 y 1.4 están intercalados. La serie no comprende $11 \times 18 \times 2 = 396$ casos porque en el periodo en cuestión hubo intervalos de tiempo sin imágenes de ecos suficientemente densos como para aplicar el método. Nótese que en la serie de más de 220 casos hay un solo caso donde el método falló. Esto fue debido a una situación muy singular donde el algoritmo de clustering extrajo una corona delimitando exactamente el borde del dominio radar. Esto fue consecuencia de una señal derivada de un proceso de empalme entre campos radar y campos del modelo necesario para aplicar una técnica de análisis bajo estudio y sin relación directa con el tema de esta comunicación. Un tal cluster es imposible desplazarlo aleatoriamente sin salirse del dominio de verificación y en consecuencia el proceso de “bootstrapping” falló.

Particularmente en la primera mitad del periodo, los p-valores obtenidos son realmente pequeños (inferiores a 5%). Aunque en algunos casos puede haber problemas de muestreo similares al comentado anteriormente (cuando la barra vertical es

corta) podemos ver que este p-valor pequeño también se obtiene en muchos otros donde tal problema no es evidente. Debemos pues concluir que HARMONIE muestra en estos experimentos una indudable pericia a la hora de ubicar las zonas de precipitación.

Para finalizar, la figura 4 representa el p-valor medio como función del alcance de la predicción (sólo las 6 primeras horas) para tres experimentos diferentes. El número de casos puede leerse en la escala de la derecha. La línea roja corresponde al control, (predicción sin dato radar), mientras que las otras tres corresponden a diferentes métodos de asimilar el dato radar. Puede verse que el impacto del dato radar es claramente positivo en las dos o tres primeras horas de la predicción. Las líneas que toman un valor por debajo del 4% en el análisis corresponden a una técnica que virtualmente suprime los errores de posición mediante un alineamiento entre “first-guess” y observaciones. Sin embargo, en estos

ensayos, el efecto positivo no muestra la permanencia deseada.

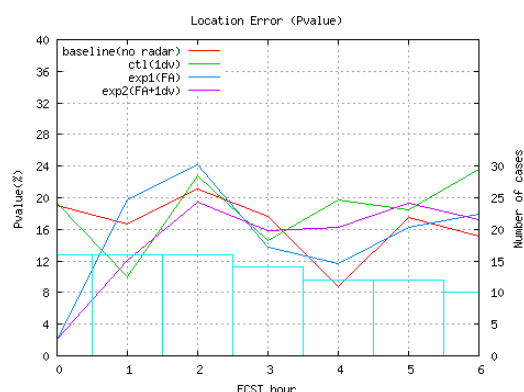


Fig 4: Aplicación del método a la verificación del impacto en la predicción del dato radar, para diferentes técnicas de asimilación del mismo.